

IMPROVING THE DATA AVAILABLE TO SIMULATION PROGRAMS

Jon W. Hand¹, Drury B. Crawley², Michael Donn², and Linda K Lawrie⁴

¹University of Strathclyde, Glasgow, Scotland

²U.S. Department of Energy, Washington, DC, USA

³Victoria University of Wellington, Wellington, New Zealand

⁴DHL Consulting, Pagosa Springs, Colorado, USA

ABSTRACT

Building performance simulation tools have significantly improved in quality and depth of analysis capability over the past thirty-five years. Yet despite these increased capabilities, simulation programs still depend on user entry for significant data about building components, loads, and other typically scheduled inputs. This often forces users to estimate values or find previously compiled sets of data for these inputs. Often there is little information about how the data were derived, what purposes it is fit for, which standards apply, uncertainty associated with each data field as well as a general description of the data.

A similar problem bedeviled access to weather data and Crawley, Hand, and Lawrie (1999) described a generalized weather data format developed for use with two energy simulation programs which has subsequently lead to a repository which is accessed by thousands of practitioners each year.

This paper describes a generalized format and data documentation for user input—whether it is building envelope components, scheduled loads, or environmental emissions—the widgets upon which all models are dependant. We present several examples of the new input data format including building envelope component, a scheduled occupant load, and environmental emissions.

INTRODUCTION

With the increasing sophistication of building performance simulation tools comes an increasing risk that the well-known Garbage-In-Garbage-Out (GIGO) phenomenon will become the equally well-known: GIGO (Garbage-In-Gospel-Out - <http://dictionary.reference.com/search?q=gigo>). Simulation programs are critically dependent on user input for a significant number of data about building components, loads, and other typically scheduled inputs. This often forces users to estimate values or find previously compiled sets of data for these inputs. Often there is little information about how the data was derived, what purposes it is fit for, which standards apply, what degree of uncertainty can be associated with each data field as well as a general description of the data. The sophistication, capability, and implied accuracy of 5 significant

decimal places of today's simulation programs easily blind users to the "garbage" data they use to describe a building.

Similar to Crawley, Hand, and Lawrie (1999) where we described a generalized weather data format developed for use with two energy simulation programs, this paper describes a generalized format and data documentation for user input—whether it is building envelope components, scheduled loads or environmental emissions. We consider our earlier effort to develop a generalized format for weather data was successful—17 major simulation tools either directly read or translate the format. And a repository of more than 800 well-documented files for location in 90 countries is now available to the simulation community for download. Yet, compiling that data set has been the work of a few people ensuring that data integrity, quality, and pedigree are well established and maintained.

We present several examples of the new input data format including building envelope component, a scheduled load, and environmental emissions. Just as access to climate data is a fit topic for the simulation community to address (Donn and Amor 1993, Crawley Hand and Lawrie 1999), the authors argue that mechanisms to hold and distribute information about the many other entities critical to the use of simulation tools are also a community issue. The goal of this approach is to develop a means of distributing simulation data not just to thermal simulation tools but to all building performance simulation tools.

The purpose of this format is to separate the simulation tool user from discussions centered on the precise numerical value of the metabolic rate for a person or the thermal capacity of granite to focus on an accurate description of the role of the person in the heating/cooling/lighting/acoustics of a building and the size and placement of the granite. Quality Assurance (Donn 1999) in building performance simulation is about ensuring that the simulation input data matches the reality in the mind of the designer. It is made easier when real-world concepts and language are to be cross-checked and compared: to deal with this directly, in this paper we illustrate our concept with reference to an example of staff in a medium sized simulation firm producing a new simulation of a primary school.

Our example firm has to deliver a performance prediction for a new school in a South Pacific climate in which they have never previously worked. The client has asked for a detailed analysis of the financial pro and cons of the design team's advice and has been known to sue consultants in the past. The project manager wants to maintain close scrutiny of the simulation model that the junior staff will produce. Obvious potential information problems are: data on the 'behavior' of a teacher; and data about granite as building cladding. Whilst checking the mathematical representation of these real world concepts is potentially more precise (5 decimal places of precision), focus on this precision will not provide any guarantee that the simulation model relates to reality.

An alternative approach to that proposed in this paper might be seen to be improved education of simulation staff. A logical end product of this professional education would be a Guild of 'approved simulationists'. The academic programs and regulation of such a Guild would be the responsibility of an organization like IBPSA. This would probably satisfy most Standards authorities whose focus is ensuring that buildings are seen to be assessed in a consistent manner. It might satisfy the project manager in our simulation firm if/s/he wanted only to be able to assign responsibility, not be involved in the guarantee of the firm's simulations. It does not however, get to the root cause of the issue – the many and varied ways in which buildings and their components can be represented mathematically.

An excellent example of the type of approach advanced in this paper can be seen in the efforts of the glazing and window industries to develop better mathematical representations of their products. In the not-too-distant past, one had to be careful from which text book one chose the shading coefficient for a sheet of glass. Some represented the 0 to 1 range of shading by glazing plus a blind system as a fraction of the total amount of solar radiation transmitted through a sheet of 3mm glass; others represented it relative to the amount transmitted through a hole in the wall. No account was taken of the incident angle of the radiation. Under these two systems the identical window with identical glazing could have two very different 'shading coefficients': 3mm glass in an otherwise unshaded window under the first system would have a shading coefficient of 1 and under the second system of around 0.9.

There are now various (e.g. WIS and WINDOW 5.2) software packages for calculating the angular dependent solar and daylight transmission as well as the R-value of whole window systems. Each is dependent on the one integrated database of optical data (International Glazing Data Library - IGDL). Simulation program users now can select a Low-E, Argon filled two-pane Insulated Glazing Unit (IGU) focusing on the accuracy of the physical description of the IGU, not on the numbers that only the physicist creating the IGDL can know well. What simulation program users want and are increasingly

able to do is export the relevant data from the window package directly into the (thermal) simulation tool.

THE NEED

What becomes clear each time one sits a group of students down to learn to use a simulation tool is that one of the biggest barriers to their learning the software is finding the right data to enter into the form in front of them. And this barrier persists as simulation is used in practice. Various approaches are taken to this: default values, libraries of 'typical' data, recommended sources / texts. Default values represent nothing real. Libraries of 'typical' data are not comprehensive and provide no clue how to add to them in a manner consistent with each different computer simulation. Textbooks provide data in a standardized format, not necessarily compatible with the format required by the computer program. Students learn quickly that it is possible to get a performance prediction from a simulation program relatively easily, but to learn to trust it takes much more time and effort.

What is true for the student is also true for our consulting firm. To be sure that they are modeling the school building's performance correctly they want to be able to check that the heat gain from the primary school teacher is modeled correctly. They want to be sure s/he is represented accurately not just in terms of metabolic rates at different times of the year, but also in terms of hours of occupancy. They want also to be sure that the granite cladding heat capacity, moisture permeability and thermal resistance are represented correctly. Their interest is in being able to guarantee to the client that these data are accurate representations of reality, not that particular numbers are precise. The origin of the numbers, the trustworthiness of their origins (their 'provenance' in art world terms) and their variability in real circumstances are essential to providing that guarantee.

With the significantly improved capability of simulation tools, the quality and depth of analysis possible has radically improved over the past thirty-five years. It should now be possible to use simulation tools for what they are best suited: examination of risk not just prediction of performance. Risk takes many forms in building performance studies. It can be the risk that the standard year used for the building performance simulation is not representative of how bad it might get if we have the hottest summer or coldest winter in 50 years. It can also be the risk that our simulation firm's primary school teacher runs the classroom in a different manner than normal – what are 'reasonable' variations from the 'typical' metabolic rates, clothing types and occupancy schedules? And, it can also be the risk that in the field the laboratory measured values of thermal performance vary due to weathering, installation or maintenance. In a more general performance

simulation the simulationist needs to be sure how critically dependent on these factors their thermal, lighting, acoustic performance predictions are.

Building energy performance prediction is increasingly being mandated by Code authorities (EU, AU, and NZ). This requires the use or development of systems for ensuring that the design tools perform adequately (ASHRAE 140 / IEA BESTEST for thermal performance; IEA Task 31/CIE TC3.33, 2005 work for lighting). For example, the NZ standards for residential (NZS 4218:2004) and non-residential (NZS4243:1996) buildings reference ASHRAE 140 / IEA BESTEST. However, to ensure that the performance predictions are indeed accurate requires not just that the software is adequate to the task, but that the users have the tools to ensure that the data they use is also adequate to the task.

The need then is for a generic building simulation database that incorporates the types of data that is used in building performance simulation, but which is independently verified and is able to be found easily. Taking the last requirement first, it is assumed that for this database to be easily found, it must be at an internet accessible address. Furthermore, as a web accessible database it most probably will be in an XML or similar tagged format. The tags not only describe the data but enable the comparison of data in a standardized manner. Using the semantic web (Berners-Lee 2001) links in their simulation tool the simulationist searches for the correct numbers to represent the thermal contribution to a school classroom of the teacher and discovers that there are three sources of this data; similarly they find there are 20 sources of physical (thermal, luminous and aural) properties of granite. However, because it is tagged, they can compare the data.

They can check for obvious differences in the teacher data such as cold climate or hot climate sources of the occupancy patterns. The data file, while machine readable by their simulation tool(s), has inbuilt tags for their benefit which clarify the source – surveys of many people, or a construct of a research body for a particular investigation. They can compare the different suppliers' data for the physical properties of granite, and particularly again the associated tags that describe its provenance: is it from an independent laboratory, in-house data produced by the quarry firm, or standard estimates from literature?

In each set of data required for a simulation there are very different standards to be applied when testing, there are very different testing methods and there are very different national and international bodies governing the reporting of the results of these tests. It is tempting to suggest that a body like IBPSA is the logical developer and maintainer of such a database. One can posit an equivalent of the International Glazing Database into which manufacturers place their data. However, who is the 'manufacturer' associated with the user data – the

teacher occupancy and metabolic information? Even more difficult to imagine is where the resources are to come from to maintain this centralized resource. When one expands beyond the very limited number of manufacturers of glazing to the realm of building materials data for thermal, visual and acoustic simulation, the task becomes impossibly large.

A more robust and easier to maintain, approach is to develop reporting procedures – formats for the data to be reported on the web so that it can be found, so that its provenance is recorded, so that its definitions are explicit in order that it is available in a form to be used by simulation programs (Donn 2004).

GENERIC SIMULATION WIDGETS

With the above rationale, the rest of this paper describes one approach to containing a diverse collection of cross referenced entities, options for containing such information and allowing access to the collection by simulationists and, potentially, by simulation tools.

The core idea is that descriptive entities used by simulation tools, although naturally diverse in nature, all have attributes which are needed to ensure their intentional and correct use.

Such attributes include:

- name in the form of a concise unique identifier (for quick retrieval in databases) and a phrase (as an aide to selection)
- documentation in the form of concise (as an aide to selection) and verbose notes
- pedigree/provenance - the source of the information, associated standards, testing procedures, date of entry etc.
- units of measure associated with each data field of an entity
- uncertainty associated with each data field of an entity

To order this potentially extensive set of lists of entities a hierarchy is useful. It partitions entities into separate file stores (e.g. acoustic properties, material properties, wind pressure coefficients). Each file store would be identified so that agents accessing entities would have access to parsing information. Separation of the files stores into separate domains of understanding facilitates the maintenance of that information in the form and with the accuracy demanded by experts in that domain.

Most entities associated with simulation fit into categories within a particular file store (e.g., for materials concrete, brick, stone, hardwoods, and softwoods). Complex entities such as environmental impacts might themselves be so diverse that categories such as maintenance on site and recycling processes will require a variety of parsing mechanisms.

Many entities associated with simulation have attributes which are, in fact the properties of other simulation entities. A classic example is a

construction, such as a wall, which is made up of an ordered list of materials with pointers to acoustic, color and environmental impact attributes. This implies a considerable potential for cross referencing within and between the various file stores and the categories and entities within them.

Broadly, entity attributes help us figure out what the entity is, where it came from, if it might be useful in the context of a particular project and what other entities it might be dependent on. Almost no simulation tool supports this level of attribution, especially over a range of simulation entities.

A SAMPLE FILE STORE

Presented below is one representation of a file store for acoustic entities. It is presented in tag-data format, but it could just as well be expressed in XML or held within relational or object oriented databases. The file store begins with a header which identifies its type (acoustic), a menu entry (Acoustic properties db) and the date it was last updated. Also in the header are units (*UNIT), sources (*SOURCE) and documentation (*DatabaseDoc). A Units line includes the tag *UNITS, a character identifier (which the data fields point to), a text label and a descriptive phrase. A Source line includes the tag *SOURCE, a character identifier and a descriptive phrase.

```
# db type, menu, date
acoustic,Acoustic db,Mon Apr 23 08:35 2003
*UNIT,-,none,no documentation
*UNIT,a,(m^2),(per m^2 of surface area)
*UNIT,b,(person),(per person)
*UNIT,c,(1000 m^3),(per 1000 m^3 volume)
*UNIT,d,(unit),(unit of element)
*SOURCE,-,no documentation
*SOURCE,b,Fasold Sonntag Winkler p.168 éd
1971
*SOURCE,c,Flumroc: Technical Doc - Lausanne
- Switzerland
*SOURCE,d,Lamoral: Probleme d'acoustique
des salles et des studios
*SOURCE,e,Isover: Technical doc - Lucens -
Switzerland
*UNCERT,-,no measurement uncertainty
supplied
# db level documentation
*DatabaseDoc
Acoustic data for a material layers or
composite constructions identified by
category type MAT or SYS Frequencies are:
100, 125, 160, 200, 250, 315, 400, 500,
630, 800, 1000, 1250, 1600, 2000, 2500,
3150, 4000.
```

Next is a header for one of the categories within the file store. Again there is an identity (MassiveMat), a type (MAT), a menu entry, a date of last modification and documentation about the category. The inclusion of a type allows agents accessing the file store to infer additional information about entities within the category. In this case the acoustic entities are for single layers only.

```
# Category,identity,type,menu,date
*CATEG,MassiveMat,MAT,Concrete & brick &
stone,Mon Apr 23 08:35:41 2003
*CategoryDoc
```

The MassiveMat category is material level properties of high mass layers.

Next are the entities (*ITEM) with concise id, menu entry, date and documentation. The references to units (per data field), source (per entity) and uncertainty (per data field) are found in the *USC line. A *USC line holds character pointers back to the tables of Units, Sources, and uncertainties defined at the start of the file store. For each data field there is a unit character, there is one source character for the entity and again for each data field there is an uncertainty character. In the example below the frequencies point to unit `a` (m²), source `b` Fasold Sonntag and uncertainty is ``. Thus uncertainty is not defined for these particular measurements. This is followed by one or more lines of data (*DAT). The type of the file store and of the category could inform the agent accessing the data how to parse the data. Another approach would be for the file store to reference an external schema or data dictionary.

```
# item id, menu, date
*ITEM,RoughCon,Rough concrete (strip.
formwork),Mon Apr 23 08:35:41 2001
*ItemDoc
Rough concrete (strip formwork) and similar
layer properties.
*USC,--aaaaaaaaaaaaaaaaa,b,-----
--- # units srcls uncert
*DAT,M,unit,0.010,0.010,0.010,0.010,0.010,0
.010,0.010,0.010,0.010,0.020,0.020,0.020,0.
030,0.030,0.030,0.030,0.030,
*ENDITEM
*ITEM,ConTile,Concrete tiling,Thu Mar 16
11:04:17 2000
*ID : Concrete tiling
*USC,--aaaaaaaaaaaaaaaaa,b,-----
--- # units srcls uncert
*DAT,C,unit,0.020,0.020,0.020,0.020,0.020,0
.020,0.030,0.030,0.030,0.040,0.040,0.040,0.
050,0.050,0.050,0.050,0.050,
*ENDITEM
*ENDCAT
*ENDDB
```

Some simulation entities can be created by pointing to entities in other file stores. A construction, such as a wall could reference color entities for each face, acoustic properties for the whole construction, environmental impacts and for each layer, entities holding information on material properties, lighting, and environmental impacts. An example of such a file store is shown below.

```
# db type, menu, date
constructions,Constructions db,Mon Feb 12
07:23:55 2001
*UNIT,-,none,no documentation
*UNIT,a,(km),(distance in km)
*UNIT,b,(m),(thickness in m)
*UNIT,c,(index),(index)
*UNIT,d,(-),(-)
*UNIT,e,(R),(gap resistance)
*SOURCE,-,from a legacy simulation database
*DatabaseDoc
Each construction entry contains the
following fields: name of matching
construction (`same` if symmetric), color
names for both faces, OPAQ:TRAN tag &
optical name, acoustic name on both faces
```

longwave emissivity on both faces, solar absorb at both faces, environmental impacts name.

For each layer: indicator for thermal or non-thermal layer, indicator for solid or air gap, material entity name, thickness (m), air gap resistance (vertical/horizontal/ sloped placement)
CAT tag,id,type,date
*CATEG,convglz,STD,EOS converted glazing items,Mon Feb 12 07:23:55 2002
*CategoryDoc
Glazing category is intended for commercial sector in Northern Europe, see convglzs for additional items.
item id, menu, date
*ITEM,dbl_glz_int,double glazed int blind,Wed Dec 20 16:18:54 2000
*ItemDoc
construction dbl_glz_int is double glazing with internal blind with optical properties from WIS analysis.
*USC,-----dddd-d---beeee-,-,-----

*DAT,same,greys50,greys50,TRAN,eos_int_win,Dgz4-12-4,Dgz4-12-4,0.900,0.900,0.500,0.500,PVCwindow,5,
*REP,t,s,CF4mm,CF_Glass,0.00600,109,0.0000,0.0000,0.0000,6mm clear float
*REP,t,g,Air,air,0.06100,0,0.1700,0.1700,0.1700,air (@ 25C)
*REP,t,s,CF4mm,CF_Glass,0.00600,109,0.0000,0.0000,0.0000,6mm clear float
*REP,t,s,Alu_spacer,Alu_spacer,0.06000,25,0.0000,0.0000,0.0000,Aluminium spacer
*REP,t,s,PVC_Frame,PVC_frame,1.00000,118,0.0000,0.0000,0.0000,PVC frame
*ENDITEM
*ENDCAT
*ENDDB

In this case, there are both static data items (*DAT), and repeating data associated with each layer (*REP). The fields in the *DAT line also include the concise identities or entities holding additional properties.

Other than differences in the contents of the *DAT and *REP lines, a range of entities can be held with an essentially common format. If all file stores follow the same convention, the bulk of the information can be accessed by relatively dumb agents. Only the decoding of the *DAT and *REP lines are dependent on the type of the file store and the category of the entity requires specific parsing and this can be delegated to specific code in a simulation tool or in a database access agent.

A similar database structure can be created for scheduled casual gains (internal process loads). An example for lighting, small power, occupancy and HVAC systems follows.

```
# db type, menu, date
schedule, Casual gains schedule db, Mon
Jan 31 11:12:35 2005
*UNIT,-,none,no documentation
*UNIT,a,(fraction),(fraction of full load)
*UNIT,b,(ON/OFF),HVAC
*SOURCE,-,no documentation
*SOURCE,a,ASHRAE Std 90.1-1989, Sect 13
*UNCERT,-,no uncertainty supplied
# db level documentation
*DatabaseDoc
Data for schedules related to casual
(internal) gains
```

Based on the schedules for performance simulation in ASHRAE Standard 90.1-1989

```
# Category,identity,type,menu,date
*CATEG,Office,Schedule, Casual Gains, Mon
Jan 31 11:12:35 2005
*CategoryDoc
A collection of lighting and receptacle
gains suitable for offices.
# item id, menu, date
*ITEM,std_off_lights,Lighting & Receptacles
office gains, Mon Jan 31 11:12:35 2005
*ItemDoc
Office lighting and small power
(receptacles) fractional schedule for 24
hour periods based on typical office
occupancy.
*USC,aaaaaaaaaaaaaaaaaaaaa,a,-----
----- # units srcs uncert
*DAT,Weekday&SummerDesignDay,unit,0.05,0.05
,0.05,0.05,0.05,0.1,0.1,0.3,0.9,0.9,0.9,0.9
,0.9,0.9,0.9,0.5,0.3,0.3,0.2,0.2,0.1,0.05,
*DAT,Saturday&WinterDesignDay,unit,0.05,0.0
5,0.05,0.05,0.05,0.05,0.1,0.1,0.3,0.3,0.3,0
.3,0.15,0.15,0.15,0.15,0.15,0.05,0.05,0.05
,0.05,0.05,0.05,0.05,
*DAT,Sunday&Holiday&AllOtherDays,unit,0.05
,0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.0
5,0.05,0.05,
0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.0
5,0.05,0.05,0.05,
*ENDITEM
*ITEM,std_off_occup, Occupancy for office
Casual Gains, Mon Jan 31 11:12:35 2005
*ItemDoc
Office occupancy fractional schedule for 24
hour periods based on typical office
occupancy.
*USC, aaaaaaaaaaaaaaaaaaaaaa,a,-----
----- # units srcs uncert
*DAT,Weekday&SummerDesignDay,unit,0.0,0.0,0
.0,0.0,0.0,0.0,0.1,0.2,0.95,0.95,0.95,0.95,
0.5,0.95,0.95,0.95,0.95,0.3,0.3,0.3,0.3,0.1
,0.1,0.05,
*DAT,Saturday&WinterDesignDay,unit,0.0,0.0,
0.0,0.0,0.0,0.0,0.1,0.1,0.3,0.3,0.3,0.3,0.1
,0.1,0.1,0.1,0.1,0.05,0.05,0.0,0.0,0.0,0.0,
0.0,
*DAT,Sunday&Holiday&AllOtherDays,unit,0.0,0
.0,0.0,0.0,0.0,0.0,0.05,0.05,0.05,0.05,0.05
,0.05,0.05,0.05,0.05,0.05,0.05,0.05,0.0,0.0
,0.0,0.0,0.0,
*ENDITEM
*ITEM,std_off_hvac,HVAC Schedule for office
Casual Gains, Mon Jan 31 11:12:35 2005
*ItemDoc
Office HVAC on/off schedule for 24 hour
periods based on typical office occupancy.
*USC,aaaaaaaaaaaaaaaaaaaaa,a,-----
----- # units srcs uncert
*DAT,Weekday&SummerDesignDay,on/off,0,0,0,0
,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,
*DAT,Saturday&WinterDesignDay,on/off,0,0,0,
0,0,0,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,
*DAT,Sunday&Holiday&AllOtherDays,on/off,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,
*ENDITEM
*ENDCAT
*ENDDB
```

And a similar one for atmospheric emissions for two energy sources:

```

# db type, menu, date
factors,Energy Use Atmospheric Emissions
Factors db,Mon Jan 31 11:12:35 2005
*UNIT,-,none,no documentation
*UNIT,a,(g/MJ),(gram of emission per MJ
energy use)
*UNIT,b,(L/MJ),(litres of water per MJ
energy use)
*SOURCE,-,no documentation
*SOURCE,a, USEPA Compilation of Air
Pollutant Emission Factors, AP-42, Fifth
Edition, Volume I: Stationary Point and
Area Sources, Supplement D, July 1998,
Chapter 1 External Combustion Sources,
Section 1.4 Natural Gas Combustion
*SOURCE,b, United States 1999 national
average electricity emissions factors based
on eGRID, 1605, AirData
*UNCERT,a,one standard deviation
*UNCERT,b,two standard deviations
*UNCERT,-,no uncertainty given
# db level documentation
*DatabaseDoc
Environmental emissions by energy source
database, data for electricity based on
local utility fuel mix. Data for other
energy sources based on published
international data
# Category,identity,type,menu,date
*CATEG,Energy_Source,Factor,Emissions, Mon
Jan 31 11:12:35 2005
*CategoryDoc
Data are in the following order: CO2, CO,
CH4, Nox, N2O, SO2, PM, PM10, PM2.5, NH3,
NMVOC, Hg, Pb, Water, Nuclear (high),
Nuclear (low).
# item id, menu, date
*ITEM,Natural_Gas,Energy Use Emissions
Factor, Mon Jan 31 11:12:35 2005
*ItemDoc
Atmospheric pollutant emissions per unit of
natural gas energy use.
*USC,aaaaaaaaaabaabaa,a,----- #
units srcs uncert
*DAT,50.23439,3.51641E-02,9.62826E-04,
4.18620E-02,9.20964E-04,2.51172E-04,
3.18151E-03,2.38613E-03,7.95378E-04,0,
2.30241E-03,1.08841E-07,2.09310E-07,
L/MJ,0,Nuclear high,g/MJ,0,0,
*ENDITEM
*ITEM,Electricity,Energy Use Emissions
Factor, Mon Jan 31 11:12:35 2005
*ItemDoc
Atmospheric pollutant emissions per unit of
electricity energy use.
USA national average
*USC aaaaaaaaaaabaabaa,b,abbccccccedee # units
srcs uncert
*DAT,168.33317,4.20616E-02,1.39858E-03,
4.10753E-01,2.41916E-03,8.65731E-01,
2.95827E-02,1.80450E-02,1.15377E-02,
1.10837E-03,3.72332E-03,3.36414E-06,
0,TBA,2.10074,0,0,
*ENDITEM
*ENDCAT
*ENDDB

```

There will be dependencies between entities in different databases/file-stores, however they are implemented. Figure 1 shows some of these. For example, *pressure distribution* entities have no dependencies and are independent of other databases/file-stores. *Optics* entities have no dependencies and are used by *constructions* entities. *Materials* entities use information from *acoustics* and *color* databases/file-stores, and *materials* entities

in turn are used by *constructions* entities. Some databases/file-stores are used by more than one higher level database. Examples of this are the *acoustics* and *environmental impacts* database/file-stores which hold data for single materials as well as properties of constructions.

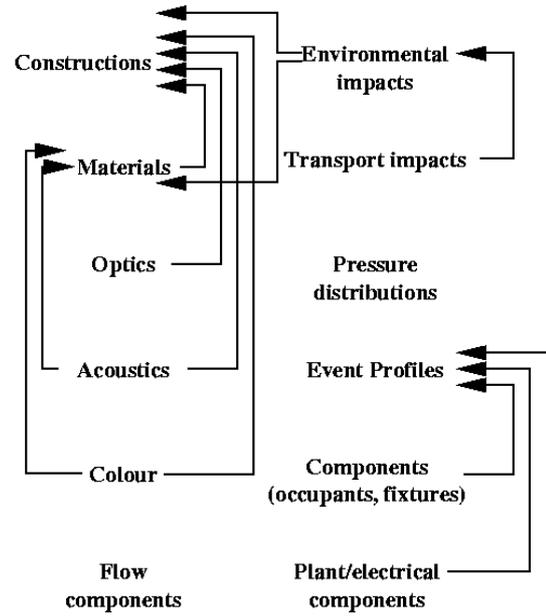


Figure 1: Relationships between databases/file-stores

IMPLEMENTATION

In the simulation firm that we described in the introduction, neither the simulationist, nor the project manager wants to deal with this level of detail in the data storage. They each have a series of questions. The simulation agent has to be able to answer these satisfactorily. The simulationist wants first to know *what weather files can I get for my South Seas island?* The web search agent in the simulation software reports that there are four files available: for Fiji and Samoa from a web site run by a government lab as part of their energy efficiency program, and one each for Nauru and for the Cook Islands on two University web sites. The agent provides links to these files, to descriptive statistics for each file, and to a 'trust score' to assist in the selection of the 'best' file.

The weather file descriptive statistics allow us to compare the average, average maxima and average minima. Armed with these averages for our site, the agent in the simulation program ranks Cook Islands then Fiji as the closest matches, both being based on larger databases. But, then the agent reports the trust score as higher for the Fiji file.

The trust score has several components. First, using the standard academic model, the Fiji file is backed by papers in two refereed journals and three significant conference proceedings, whereas the Cooks file has references only to a PhD thesis and

one minor conference paper. The second part of the trust score is a 'self-interest level' rating – the more the provision of the data is likely to promote a particular product for sale, the lower its reliability. In this case the providers of both the Fiji and the Cooks files have nothing to gain from provision of the data so have the same rating. The third part of the trust score applies the 'ebay principle': users' ratings of the data. Surprisingly, the rating of the Fiji data from the government lab is a little lower than the Cook Island file; the simulation program agent reports that the government lab has recently adopted a policy of charging a fee for each computer that the climate data is installed on.

The simulation firm has three other projects to analyse that are in the same climate as the proposed building. As a result, the simulationist recommends to the project manager that the Fiji data is used.

Now the simulationist's question posed to the simulation program agent is *what is a reasonable model of a primary school?* First the agent provides a basic structured simulation model of a school. This is part of the sample set of buildings provided with the simulation program, but the XML tags within its data structure provide a full provenance. It is in fact a description of a School recently constructed in Auckland, New Zealand. The agent provides a trust score for the file which is high in school type match, but low in climate match: Auckland at 35S Latitude, is clearly 'South Seas' but too far South. The simulationist browses the rest of the file examining the provenance carefully: the file is from a reputable research organization (based on its publication record); the file refers to a journal paper where the monitored energy use is reported. All contributing to an overall high score, but eventually he decides that the climate differences are so great that he will start to build the simulation model from scratch.

The agent plays no role in the entering of the dimensions of the classrooms. However, as the data entry continues the question of where to find quality data becomes critical. The simulationist now wants to know: *First, what is the model to be used of the hours of use of the classrooms?* How many people are in a classroom and what is their direct and latent heat contribution to the space? To complicate matters, the architect has read a theory that high mass buildings with good ventilation are a viable design option in this hot, humid climate. She has a composite concrete construction which uses outside facing panels of granite. The question the project manager poses to the agent is *how to be sure that the data used to describe this sandwich of concrete, insulation and granite is modeled with sufficient accuracy that he can produce persuasive evidence to back the simulationist's design recommendations in the next design team meeting?*

The agent reports 3 sources of data readily available on the web for use patterns in schools. On the basis merely of the descriptive statistics included in the XML tag description of the data, only one looks

useful. The average number of students in the other two files is very different from that planned for this school. They both seem to be for 'open-plan' schools from the 1970's with several classes in the one room. Fortunately, the trust score for the one apparently relevant source is high: it is backed up by documentation comprising a paper in a refereed conference proceedings; there is again no marketing advantage accruing to the provider from the simulationist's use of the data; and 18 different people have reported favourably on their use of the data (though 15 are apparently from the same university simulation class!).

Finally, the agent provides 20 links to the thermal properties for granite. These are classified in the descriptive statistics slightly differently. The agent can average the R-values across all 20 and rate each one as to how far from the average it sits; because the data is in a mixture of units, the agent has read the XML units tag and has converted them all to the same SI units R-value form that are standard for the simulationist's office, but it labels them as to their origin (k-value, conductivity, resistance, SI, non-standard Metric, Imperial etc). The trust score is now more complex: very few data points are backed with published papers; several are published generic values from tables published by bodies like ASHRAE or CIBSE; some are from laboratory tests. For all the reports the XML descriptive tag for the material is highly important because the many types of granite are described here in ways that can be communicated directly to the architect so she can consult her supplier catalogue and determine the closest match. Within these 5 close matches, only two are published with the associated laboratory tests linked. This provides a high trust score on academic credibility, but then only one of these two is from an independent source; one is from an in-house laboratory for a company marketing granite as a building material. Finally, the user ratings for these last two laboratory tested data sources are approximately equal. The simulationist informs the simulation agent that the model will use the independent laboratory's data but also notes that it should record for parametric testing the in-house laboratory data as this may well provide an indication of the range of likely values for the granite thermal properties and the agent should report in the output of the simulation what influence choosing one of these numbers has had on the simulation.

CONCLUSION

The end result for our example firm of simulationists is a thermal model that is able to be shown (in court or in a client meeting!) to be well-documented. The audit trail to the documentation is automatically recorded by the simulation agent and is able to be printed out separately as and when needed. It is recorded in the XML format of the output of the simulation, as just another tag labeling the simulation quality. And, perhaps most significantly the simulation has an associated trust score based in

part on the accumulated trust scores of the input data. To gain the highest trust score on re-publication in an XML searchable form on the web, this simulation input file with its associated output must also pass the same trust test as the input file: good academic credibility of the author organization; no benefit accruing to the provider of the files (sample files provided by vendors of simulation programs would be less trustable than those from independent organizations); and users of the data would record their confidence in it.

The key to this system is that it uses the strengths of the internet: it is a distributed system, no one organization is responsible for entering and checking the information; it is able to be incrementally implemented immediately because it does not require a complete re-write of systems and databases, since the nature of the database format of input files for most simulation programs mean that they can be readily converted to the XML self-annotating format of web databases. Also, the system can be applied to lighting, acoustic and air flow simulations as readily as it can be to thermal simulations.

The flaw that is normally suggested for the system centers on the trust score. Many people find it hard to accept that the system might be able to be distributed. It is difficult for them to imagine a self-policing process of quality control – of trust scores – for this data. Placing one’s simulation data integrity in the hands of malicious grad students is a common fear. However, as proposed, the assessment of this risk is a part of the development of the trust score. Those who provide data that is false face the same very public disenfranchisement that deters people from falsifying data for submission to a refereed journal or conference. The trust score is the core of this process. It is the development of the trust score process and the specification of the XML tags with which this data exchange occurs that is properly the work of an organization like IBPSA.

REFERENCES

- ASHRAE. 2004. *ANSI/ASHRAE Standard 140-2004*, “Standard Method of Test for Evaluation of Building Energy Analysis Computer Programs,” September 2004. Atlanta: ASHRAE.
- Berners-Lee. 1999. *Weaving the Web – the original design and ultimate destiny of the world wide web by its inventor*. San Francisco: Harper.
- Crawley, D.B., J.W. Hand, L.K. Lawrie. 1999. “Improving The Weather Information Available To Simulation Programs,” in *Proceedings of Building Simulation '99*, Volume 2, pp. 529-536. Kyoto, Japan, 13-15 September 1999. IBPSA.
- Donn, Michael. 2004 “Simulation of Imagined Realities - environmental design decision support tools in architecture”. *PhD Thesis*, Victoria University, School of Architecture, Wellington, NZ.
- Donn, Michael. 1999. “Quality Assurance—Simulation and the Real World,” in *Proceedings of Building Simulation '99*, Volume 3, pp. 1139-1146. Kyoto, Japan, 13-15 September 1999. IBPSA.
- Donn, M., R. Amor. 1993. “The Creation of Weather Data Files to Fulfill the Simulation Needs of New Zealand,” in *Proceedings of Building Simulation '93*, pp. 495-499. Adelaide, Australia. 16-18 August 1993. IBPSA.
- CIE TC3.33 2005. Technical Report: Test Cases To Assess The Accuracy Of Lighting Computer Programs, *CIE Technical Committee 3.33*.
- NZS 4218. 2004. *Energy Efficiency—Housing and Small Building Envelope*. Wellington: Standards New Zealand.
- NZS 4243. 1996. *Energy Efficiency—Large Building Envelope*. Wellington: Standards New Zealand.
- <http://dictionary.reference.com/search?q=gigo> – last accessed, Feb 2005.